

RESEARCH

Open Access



Integrating plasma protein-centric multi-omics to identify potential therapeutic targets for pancreatic cancer

Siyu Zhou^{1†}, Baian Tao^{1†}, Yujie Guo^{1†}, Jichun Gu¹, Hengchao Li¹, Caifeng Zou¹, Sichong Tang², Shuheng Jiang^{3*}, Deliang Fu^{1*} and Ji Li^{1*}

Abstract

Background Deciphering the role of plasma proteins in pancreatic cancer (PC) susceptibility can aid in identifying novel targets for diagnosis and treatment.

Methods We examined the relationship between genetically determined levels of plasma proteins and PC through a systemic proteome-wide Mendelian randomization (MR) analysis utilizing cis-pQTLs from multiple centers. Rigorous sensitivity analyses, colocalization, reverse MR, replications with varying instrumental variable selections and additional datasets, as well as subsequent meta-analysis, were utilized to confirm the robustness of significant findings. The causative effect of corresponding protein-coding genes' expression and their expression pattern in single-cell types were then investigated. Enrichment analysis, between-protein interaction and causation, knock-out mice models, and mediation analysis with established PC risk factors were applied to indicate the pathogenetic pathways. These candidate targets were ultimately prioritized upon druggability and potential side effects predicted by a phenome-wide MR.

Results Twenty-one PC-related circulating proteins were identified in the exploratory phase with no evidence for horizontal pleiotropy or reverse causation. Of these, 11 were confirmed in a meta-analysis integrating external validations. The causality at a transcription level was repeated for neutrophil elastase, hydroxyacylglutathione hydrolase, lipase member N, protein disulfide-isomerase A5, xyloside xylosyltransferase 1. The carbohydrate sulfotransferase 11 and histo-blood group ABO system transferase exhibited high-support genetic colocalization evidence and were found to affect PC carcinogenesis partially through modulating body mass index and type 2 diabetes, respectively. Approved drugs have been established for eight candidate targets, which could potentially be repurposed for PC therapies. The phenome-wide investigation revealed 12 proteins associated with 51 non-PC traits, and interference on protein disulfide-isomerase A5 and cystatin-D would increase the risk of other malignancies.

[†]Siyu Zhou, Baian Tao, and Yujie Guo contributed equally to this study.

*Correspondence:

Shuheng Jiang

shjiang@shsci.org

Deliang Fu

surgeonfu@163.com

Ji Li

lijli@huashan.org.cn

Full list of author information is available at the end of the article



Conclusions By employing comprehensive methodologies, this study demonstrated a genetic predisposition linking 21 circulating proteins to PC risk. Our findings shed new light on the PC etiology and highlighted potential targets as priorities for future efforts in early diagnosis and therapeutic strategies of PC.

Keywords Pancreatic cancer, Plasma proteome, Therapeutic target, Mendelian randomization

Introduction

Pancreatic cancer (PC) is one of the leading causes to cancer death worldwide with increasing incidence and a meager 5-year survival rate of less than 9% [1, 2]. Approximately 80% of patients present with advanced and unresectable disease at diagnosis, which is partially attributed to the asymptomatic nature and difficulty in early detection [3]. Precancerous or early-stage lesions cannot be efficiently recognized merely by imaging alterations, implying the significance of exploring reliable diagnostic biomarkers [4]. But even for resectable PC, the prognosis of patients is optimistic as a result of rapid postoperation relapse and chemotherapy resistance [5]. Thus, novel available therapeutic strategies are warranted.

The plasma proteins as vital components in circulating blood, produced by cellular leakage and active secretion, are involved in various crucial physiological and pathological processes, and can thereby act as a reflection of the overall physical condition as well as possible druggable targets for illnesses [6–8]. Specifically, several circulating proteins are suggested to be biomarkers for inflammation, infection, and some systemic diseases [9–11]. With regard to malignancies, a number of cross-sectional studies have looked into the discrepancy in circulating protein levels between cancer sufferers and healthy controls in an attempt to establish the intricate protein-carcinogenesis connection [12–15]. But the nature of their observational studies restricts the reliability of conclusions due to potential confounding bias and reverse causation [16].

Recently, a series of large-scale proteomic research have identified the protein quantitative trait loci (pQTLs), enabling the causality inference for the effect of plasma protein on PC susceptibility via a two-sample Mendelian Randomization (MR) method, which utilizes genetic variants as instrumental variables to mimic randomized controlled trials [17–19]. Since MR results are less likely to be biased by confounders and reverse causation, the MR method is widely applied in investigating the causative factors for outcomes, such as the causal correlation of peripheral metabolites or gut microbiome with PC [20–22]. Of note, as the extension of MR methodology, proteome-wide MR studies focus on the genetic-determined circulating protein concentration and disease etiology and have been employed

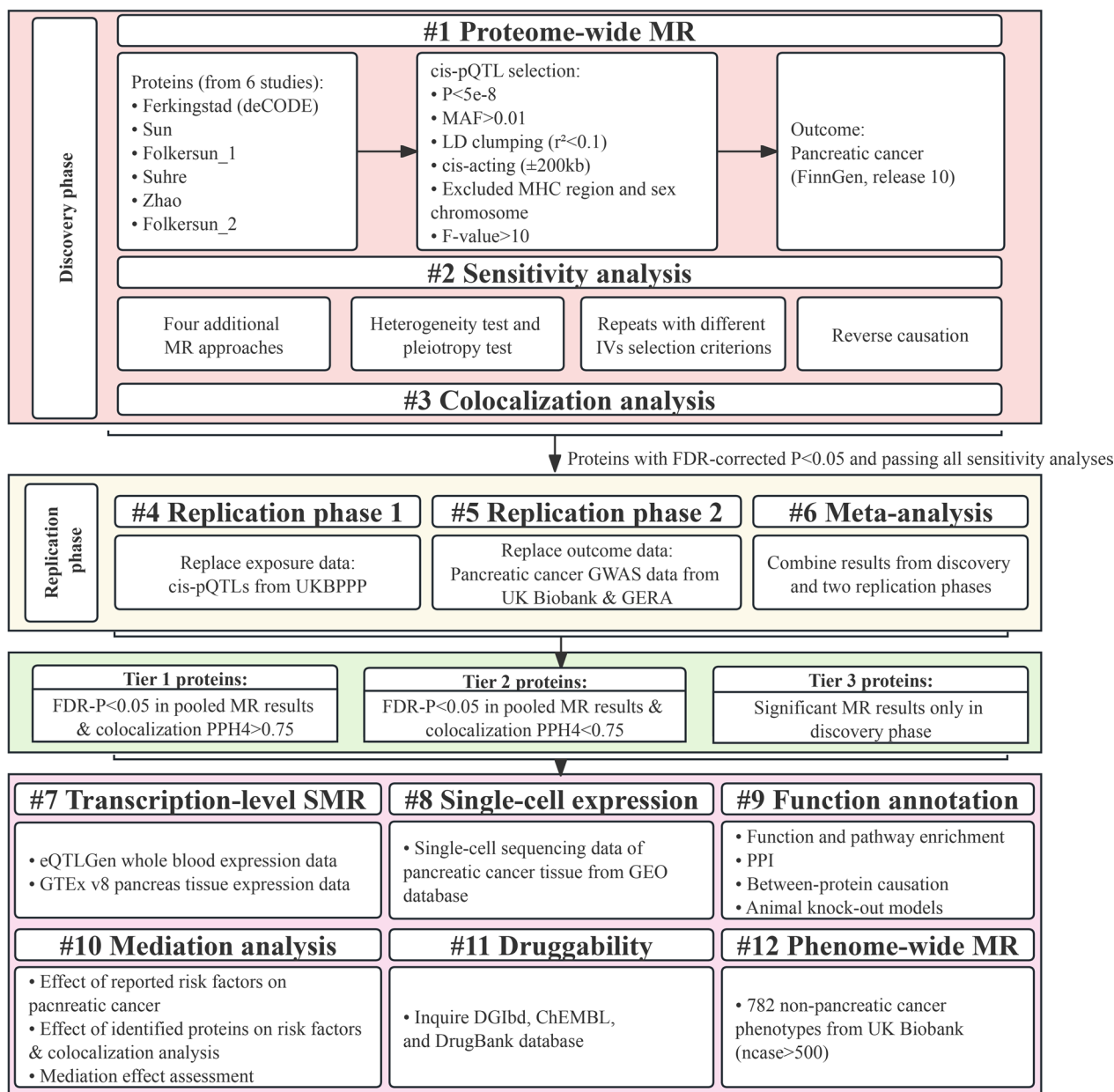
for exploring carcinoma-related biomarkers or promising interference targets in tumors like colorectal cancer, breast cancer, and lung cancer [23–26].

In the present study, we integrated cis-pQTL data for a proteome-wide MR analysis to identify PC-associated plasma proteins. Bidirectional MR, replicative validation and meta-analysis, and Bayesian colocalization were used to confirm the primary results. Then the corresponding protein-coding gene expressions were also analyzed regarding their causal effect on PC and their expression pattern in single-cell types. The function and involved pathways of these targets were preliminarily investigated through enrichment analysis, knock-out mice models, and between-protein interaction and causation. The interplay network between circulating proteins, known PC risk factors, and PC was further analyzed and discussed. Finally, drug-target databases were inquired to prioritize the druggable targets, and a phenome-wide MR was conducted to evaluate the drug safety and repurposing.

Methods

Overall design

The workflow and methodology of this study are outlined in Fig. 1. In brief, cis-pQTL data derived from six publicly accessible datasets were employed to conduct a proteome-wide, two-sample MR in the primary study phase. Subsequently, a three-part analytic protocol was applied to enhance and expand our initial findings. For part one, we employed several sensitivity analyses, bidirectional MR, Bayesian colocalization, external replications and meta-analysis, and replicative MR analysis in transcription level to validate the primary proteome-wide MR results. For part two, single-cell type expression analysis, GO/KEGG enrichment, mutual causality, protein–protein interaction (PPI) network, single gene knock-out mice models, and mediation analysis, were used to annotate the function and infer the potential pathogenic pathways of these PC-associated candidates obtained from analysis in part one. For part three, the druggability and possible side effects estimated by a phenome-wide MR were assessed for prioritizing these therapeutic targets. The P-values were all adjusted by the Benjamini–Hochberg false discovery rate (FDR) in multiple tests in our study.



Part 1: analyses #1-7, for principal discovery and multi-dimension validation;

Part 2: analyses # 8-10, for inference of the function and potential pathogenic mechanism of candidate proteins in part 1;

Part 3: analyses #11-12, for druggability and drug side-effect assessment for candidate proteins in part 1.

Fig. 1 Flowchart of the study design. MR, Mendelian randomization; IV, instrumental variable; pQTL, protein quantitative trait loci; FDR, false discovery rate; UKBPPP, UK Biobank Pharma Proteomics Project; GWAS, genome-wide association study; GERA, Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging cohort; GTEx v8, Genotype-Tissue Expression Project version 8; GEO, Gene Expression Omnibus; PPI, protein-protein interaction; DGIdb, Drug Gene Interaction Database

Study datasets and genetic instruments selection

In the exploratory phase, genome-wide association study (GWAS) summary statistics regarding PC were acquired from the FinnGen consortium R10 release (<https://www.finnngen.fi/en>) [27]. In the present research, PC was

defined as malignant neoplasm of the pancreas, incorporating pancreatic ductal adenocarcinoma and other pathological types of malignant pancreatic tumors. Summary statistics of genetic associations with circulating proteins were obtained from six distinct large-scale proteomic

studies (Ferkingstad et al., 4907 proteins; Sun et al., 3282 proteins; Folkersen_1., 82 proteins; Suhre., 1124 proteins; Folkersen_2., 90 proteins; Zhao., 91 proteins [6, 28–32]). Detailed descriptions of the above datasets can be found in original publications. We harmonized these proteomic data by remapping protein IDs onto corresponding gene symbols. For proteins presented in multiple datasets, or those with different probes or isoforms, we calculated the proportion of variability (R^2) (see below), and the one with the largest R^2 was retained. To satisfy the basic assumptions of MR, we filtered the extracted pQTLs upon the following criterion: (1) pQTLs with genome-wide significant ($P < 5E-8$) association with any protein; (2) minor allele frequency (MAF) > 0.01 ; (3) linkage disequilibrium (LD) R^2 between SNPs was controlled < 0.1 within 10 Mb; (4) cis-pQTL was defined as pQTL being cis-acting (within 200 kb upstream or downstream of the protein-coding gene region); (5) SNPs located in the major histocompatibility complex (MHC) region (chromosome 6, 31–33 Mb) and sex chromosome were excluded; (6) R^2 and F-value were computed to evaluate the strength of instrumental variables (IVs) ($R^2 = 2 * \text{MAF} * (1 - \text{MAF}) * \beta * \beta$; $F\text{-value} = R^2 * (N - 2) / (1 - R^2)$) [33], and SNPs with $F\text{-value} < 10$ were removed; (7) cis-pQTL should not be directly associated with PC ($P\text{-value} > 1E-5$). The involved GWAS studies mainly enrolled participants of European ancestry and had all received approval from their corresponding ethical review committees.

Proteome-wide MR, sensitivity analysis, and reverse MR analysis

The “TwoSampleMR” R package was utilized for a proteome-wide MR analysis [34]. The MR methodology employed SNPs as IVs to infer causal relationship between two traits, and the Wald ratio ($\text{No.SNPs} = 1$) and inverse variance-weighted (IVW) algorithms ($\text{No.SNPs} > 1$) were applied as principal MR approaches since they were most efficient when all IVs were valid since [35]. The Wald ratio algorithm calculated the effect ratio of one variant in exposure and outcome. When there was at least two instruments, the IVW algorithm was used to combine the ratio estimates of each variant in a meta-analysis model. Before MR analysis was performed, we harmonized the exposure and outcome data using the “harmonise_data” function. This process extracted IVs that overlapped in the filtered exposure data with the outcome data and automatically removed incompatible and palindromic SNPs. The presence of heterogeneity was assessed using Cochran’s Q test, and a test P-value less than 0.05 indicated heterogeneous IVs. In this case, a random-effect IVW model would be used. Otherwise, a fixed-effect IVW MR was performed. The P-values of the

proteome-wide MR results were corrected with the Benjamini–Hochberg FDR method, and causal associations with FDR-corrected P-values less than 0.05 were considered significant. Additionally, the MR-Egger regression intercept test, MR-Pleiotropy Residual Sum and Outlier (MR-PRESSO) methodology, and MR-PRESSO global test were employed to evaluate the horizontal pleiotropy [36, 37]. Subsequently, MR analyses with four additional approaches including weighted median, MR-Egger, weighted mode, and simple mode were performed as part of sensitivity analyses. Since the MR results were potentially susceptible to the IVs selection, we re-analyze the data after modifying the IVs inclusion criterion by taking the parameter of LD R^2 threshold of 0.001, 0.01, 0.2, and 0.3, respectively. Additionally, the presence of reverse causation was assessed by an inverse MR. However, only two SNPs were initially extracted as IVs proxied for PC after pruning instruments with a stringent threshold for P-value ($P\text{-value} < 5E-8$) and LD clumping ($r^2 = 0.001$). Thus, a broader threshold for the P-value of $5E-6$ was adopted as a replicative validation in the reverse MR. Upon integrating results of all above sensitivity analyses, potential targets with robust evidence in the discovery phase were defined as: (1) significantly associated with PC after multiple tests correction; (2) no pleiotropic outliers detected, or significant MR results in the re-analysis after removing outliers; (3) absence of horizontal pleiotropy revealed by Egger intercept or MR-PRESSO global test; (4) identical effect direction to primary results in sensitivity analyses using additional MR approaches and varying LD parameters; (5) no evidence showing reverse causation.

Bayesian colocalization

Circulating proteins in significant relation to PC and passing all sensitivity tests were analyzed with Bayesian colocalization using “coloc” R package to illuminate whether a protein and PC were linked to a shared causal variant or the association was driven by the confounding of linkage disequilibrium [38]. Aligning with previous studies, we adopted default parameters of $p1 = 1E-4$, $p2 = 1E-4$, and $p12 = 1E-5$ in this process [39]. $p1$ and $p2$ represented the prior probability that a SNP is significantly correlated with protein and PC risk, respectively, and $p12$ represented the prior probability of a SNP being associated with both traits. For each locus, the posterior probabilities of the following five hypotheses were assessed: (1) $H0$: no causal variant for either plasma protein or PC; (2) $H1$: one causal variant only for protein; (3) $H2$: one causal variant only for PC; (4) $H3$: two different causal variants for protein and PC respectively; (5) $H4$: one shared causal variant for both PC and plasma protein. High-support evidence of colocalization was

considered in cases with the posterior probabilities of H4 (PPH4) over 0.75, and medium-support evidence of colocalization was defined as PPH4 less than 0.75 but greater than 0.5 [40].

Replication and meta-analysis

We repeated our primary analysis in a two-phase validation. In replication phase 1, cis-pQTLs were obtained from a plasma proteomic association study in UK Biobank [41], which incorporated 2923 proteins and 54,219 European ancestry individuals, while in replication phase 2, the genome-wide association data for PC was replaced by an integrated GWAS study involving population from both UK Biobank and the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging cohort (GERA) with a sample size of 411,013 [42]. All of the above datasets were essentially GWAS data of the same phenotype with that of the exposure or outcome in the discovery phase. They had adequate sample size and large number of measured SNPs. These GWAS studies all incorporated European populations, and there was no population overlap in each exposure-outcome pair because they were from distinct research cohorts. So the selection for above validation datasets met the requirements of MR analysis and ensured these data were suitable to be employed in further replicative phases. The cis-pQTL filtering process was identical to that of the discovery phase, and sensitivity analyses (heterogeneity test, pleiotropy test, and supplementary MR approaches) were performed as usual. Moreover, to expand our findings, proteins without significant links to PC in the primary stage were also analyzed using alternative data sources. Finally, a meta-analysis was conducted to combine MR estimates from the discovery phase and two replication phases, which would be deemed as the ultimate results of external validation. Heterogeneity of meta-analysis was assessed by the statistics I² to determine the use of random effect or fixed effect models [43]. We then categorized the identified PC-related candidates that passed all sensitivity analyses in the primary stage into three tiers according to the evidentiary strength of colocalization and external validation: (1) tier 1 proteins: FDR-corrected P-value < 0.05 in meta-analysis, consistent effect direction in discovery and validation phases, and colocalization PPH4 > 0.75; (2) tier 2 proteins: FDR-corrected P-value < 0.05 in meta-analysis, consistent effect direction in discovery and validation phases, and colocalization PPH4 < 0.75; (3) tier 3 proteins: unsuccessful replication in external validation.

Transcriptome-level MR and SMR analysis

For the sake of further investigating the causation of the corresponding protein-coding genes' expression on

PC, we obtained full expression quantitative trait loci (eQTL) data for whole blood tissue from the eQTLGen Consortium (<https://eqtlgen.org/>), which comprised genetic associations with the expression of 16,987 genes among 31,684 mostly healthy participants [44]. The selection standard for cis-eQTLs was the same as that of cis-pQTLs (see above), and the acquired cis-eQTLs for candidate targets were then employed in the subsequent transcription-level two-sample MR analysis. In addition, the summary-data-based MR (SMR) test using the top hit eQTL as instrument was also implemented with SMR software (SMR v1.3.1) as a sensitivity analysis. And the heterogeneity in independent instrument (HEIDI) was conducted to distinguish the identified relationships from pleiotropy and genetic linkage [45]. The SMR-formatted cis-eQTLs data could be accessed from publicly available link (https://molgenis26.gcc.rug.nl/downloads/eqtlgen/cis-eqtl/SMR_formatted/cis-eQTL-SMR_20191212.tar.gz). Likewise, the SMR and HEIDI tests were also performed to explore the gene-PC association in pancreas tissue by utilizing the SMR-formatted cis-eQTLs data acquired from <https://yanglab.westlake.edu.cn/software/smr/#DataResource>. Results would be considered positive and valid when the P-value for SMR was less than 0.05 and the P-value for HEIDI test was over 0.05.

Single cell-type expression analysis

We downloaded the single-cell RNA sequencing (scRNA-seq) data of target protein-coding genes in 16 PC samples from the Gene Expression Omnibus (GEO) database (Registration number: GSE155698) [46]. The scRNA-seq data was then processed with "Seurat" R package [47]. The created and merged Seurat object incorporated 49,333 cells and 32,738 features. Firstly, in order to obtain high-quality data on single-cell RNA expression, the following filtering standard was established: 1. exclusion of genes with expression lower than five counts in one cell; 2. exclusion of cells with < 300 or > 4000 measured genes; 3. exclusion of cells with > 10% mitochondrial contamination. As a consequence, a total of 28,994 high-quality cells and 23,384 features were remaining for further analysis. Data after quality control would be normalized with "NormalizeData" function, which was used to normalize raw read counts by applying a scale factor of 10,000 and logarithmically transforming the values to stabilize the variance across genes with different expression levels. Subsequently the "SingleR" R package was employed for cell clusters annotation, which assigned cell identities by correlating single-cell RNA expression profiles with reference datasets of known cell types, enabling precise identification and analysis of distinct cell populations [48]. To illustrate whether the expression of the target gene was enriched in a specific cell cluster, differential

gene expression across cell types was analyzed using the Wilcoxon Rank Sum test, and the enrichment would be defined as significant when FDR-corrected P-value < 0.05 and $|\text{Log}_2(\text{fold-change})| > 1$.

Function and pathway enrichment

For exploring the potential biological implication of identified PC-related circulating proteins, enrichment analysis with regard to the GO function terms (biological processes, cellular components, and molecular functions) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was applied [49]. This enrichment methodology was conducted by evaluating whether the presence of specific genes within a given pathway significantly exceeded what would be anticipated by chance, as determined by the proportion of genes in the background dataset associated with that pathway. Functions or pathways with FDR-corrected P-value less than 0.05 were deemed significantly enriched.

Protein–protein interaction (PPI) and mutual causation

The protein–protein interaction (PPI) network was constructed with the Search Tool for the Retrieval of Interacting Genes (STRING) database (<https://string-db.org>) with a minimum interaction confidence score of 0.4. To further investigate the interplay between circulating protein levels, we carried out a series of two-sample MR analyses pair by pair of those candidate target proteins with each other.

Animal knock-out models

For circulating proteins shown as potential therapeutic targets, we queried the single gene knock-out mice models through the Mouse Genome Informatics (MGI) website (<http://www.informatics.jax.org>) to verify their biological function, as well as the possible side effects that might be brought about by the targeted therapy. Phenotypes in relation to single gene knock-out were manually classified and displayed in three categories: neoplastic phenotypes, phenotypes of the digestive system, and phenotypes of others.

Mediation analysis

The etiology of malignant tumors was complicated, and the effect of proteins on tumorigenesis might not be direct but indirect through established risk factors. In order to validate the above hypothesis, we designed a four-step analysis protocol: (1) identify risk factors for PC through previous literature; (2) verify the causation of these risk factors on PC via a two-sample MR (Effect1); (3) compute the MR estimate of PC-associated circulating proteins on these risk factors (Effect2) and conduct colocalization analysis; (4) estimate the mediation effect

and direct effect. The total effect from protein to PC was equal to the MR estimate obtained from the proteome-wide MR analysis of the discovery phase, and the mediation/indirect effect was calculated as $(\text{Effect1} * \text{Effect2})$, while the direct effect was calculated as $(\text{Effect}_{\text{total}} - \text{Effect}_{\text{mediation}})$ [50]. There would be a possibility that the mediation effect existed if the effect direction of all causal pairs in the association between proteins, PC, and risk factors followed right logic. The confidence interval (CI) and P-value of mediating effect were estimated by the delta method.

Druggability assessment and phenome-wide MR analysis

Three drug-targets databases (Drugbank: <https://go.drugbank.com/>; ChEMBL: <https://www.ebi.ac.uk/chembl/>; DGIdb: <https://www.dgiddb.org/>) were queried to identify available drugs targeting the candidate circulating proteins. Details of the drugs and drug-gene interaction were documented. Furthermore, a phenome-wide MR analysis was performed to appraise the drug safety and repurposing. Summary statistics of genetic association with extensive phenotypes were collected from UK Biobank (<https://pheweb.org/UKB-SAIGE/>), and only traits with cases over 500 were retained as outcomes in the phenome-wide MR. All the cis-pQTLs proxied for plasma proteins and all the parameters used in MR process were identical to that of the discovery phase.

Results

Proteome-wide MR identified 21 plasma proteins causally affecting PC susceptibility in the discovery phase

In the discovery phase, after removing SNPs with P-value > 5E-8 or MAF < 0.01, IVs were available for 4,790, 2,229, 63, 356, 85, and 75 proteins in the proteomic studies of Ferkingstad et al., Sun et al., Folkersen_1., Suhre., Folkersen_2., and Zhao., respectively. After LD clumping and removal of pQTLs located in sex chromosome or MHC region and pQTLs located away from protein-coding gene region, 1,774, 658, 32, 228, 72, and 61 proteins were retained respectively. After deletion of IVs with F-value < 10 and harmonization, a total of 28,050 SNPs were available to proxy 2,781 proteins. Then, proteins that appeared in more than one dataset were eliminated and the one with the largest R² sum was retained. Finally, a total of 19,379 SNPs were applied as IVs for 1,751 proteins in the primary MR analysis. The median number of cis-pQTLs used for proxying single protein was six (ranging from one to 90). And instruments for most of the proteins (1372/1751) were derived from the study of Ferkingstad et al. [28], according to their R² sum.

The primary results revealed 21 significant protein-PC association pairs after correcting P-values with FDR (FDR-corrected P-value < 0.05), as displayed in Fig. 2

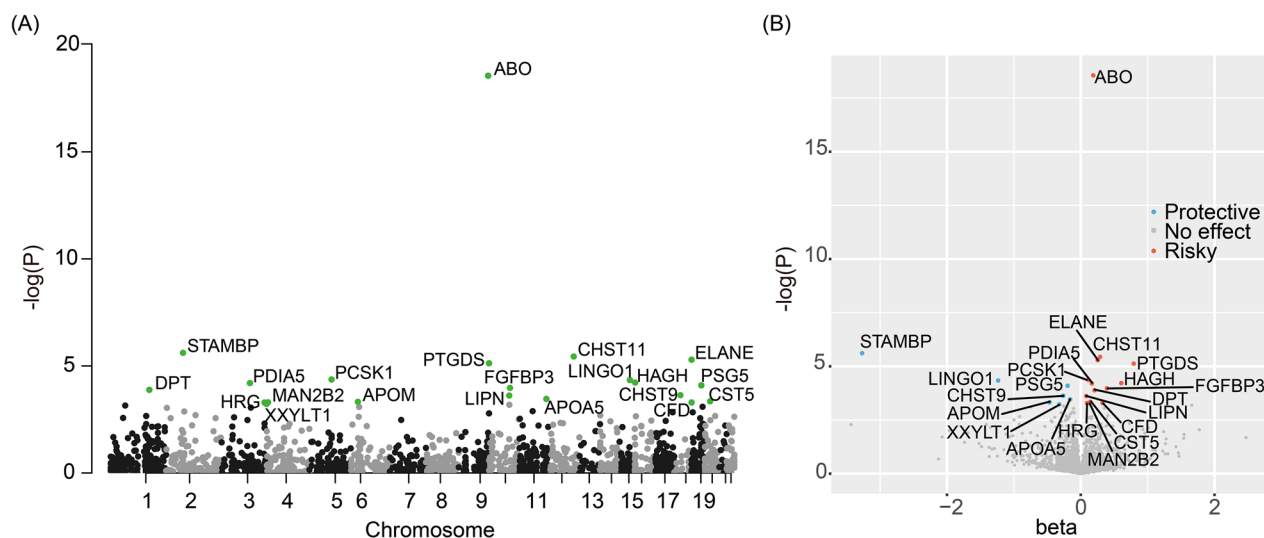


Fig. 2 The Manhattan plot (A) and volcano plot (B) of the proteome-wide Mendelian randomization analysis in the discovery phase

in the form of a Manhattan map and volcano plot. In the sensitivity analysis with additional MR approaches (MR Egger, weighted median, weighted mode, and simple mode), the direction of causality of significant findings were all in concordance with the primary conclusions (Additional file1: Table S1). In addition, no inter-SNPs heterogeneity was found via Cochran's Q test, and no horizontal pleiotropy was observed through the Egger intercept test and the MR-PRESSO global test for these positive MR results. Then we altered the clumping parameter ($LD R^2=0.001, 0.01, 0.2, \text{ and } 0.3$, respectively) in four separate repeats for the 21 identified proteins. Of the 84 replicative inspections, none of them demonstrated an effect direction contrary to the original estimates, and the majority of the MR analyses (74/84) still yield a statistically significant association (Additional file1: Table S2). In terms of the reverse causation, no obvious impact from PC on circulating protein concentrations was found after multiple testing corrections, whenever the IVs for PC were selected based on a strict ($5E-8$) or relatively loosen ($5E-6$) P-value threshold (Additional file1: Table S3). Collectively, the associations between PC and 21 genetically predicted protein levels passed all of the sensitivity analyses, implicating robust evidence.

Of the 21 PC-associated circulating proteins, 14 proteins exhibited tumor-promoting efficacy, including ABO (Histo-blood group ABO system transferase), PTGDS (Prostaglandin-H2 D-isomerase), CFD (Complement factor D), CHST11 (Carbohydrate sulfotransferase 11), ELANE (Neutrophil elastase), HARG (Hydroxyacyl-glutathione hydrolase, mitochondrial), CST5 (Cystatin-D), PCSK1 (Neuroendocrine convertase 1), HRG

(Histidine-rich glycoprotein), LIPN (Lipase member N), MAN2B2 (Epididymis-specific alpha-mannosidase), DPT (Dermatopontin), PDIA5 (Protein disulfide-isomerase A5), and FGFBP3 (Fibroblast growth factor-binding protein 3). Among these candidates, ABO yielded the most prominent causality (OR (95%CI)=1.21 (1.16–1.26), FDR-corrected P-value=4.88E-16).

The other 7 plasma proteins displayed a protective action against PC onset, including CHST9 (Carbohydrate sulfotransferase 9), PSG5 (Pregnancy-specific beta-1-glycoprotein 5), STAMBP (STAM-binding protein), LINGO1 (Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1), APOA5 (Apolipoprotein A-V), APOM (Apolipoprotein M), and XXYLT1 (Xyloside xylosyltransferase 1), among which the STAMBP showed the most significant negative causation (OR (95%CI)=0.04 (0.01–0.15), FDR-corrected P-value=2.12E-3).

Colocalization analysis

To distinguish the causal relationship between genetically determined circulating protein levels and PC from linkage disequilibrium, the colocalization analysis was applied. Among the 21 candidate proteins, the PPH4 values for ABO (PPH4=0.967) and CHST11 (PPH4=0.757) were over 0.75, suggesting high-support evidence for an association linked to a shared causal variant (Additional file1: Table S4). Besides, medium-level evidence of colocalization (PPH4>0.5) was observed linking LINGO1 (PPH4=0.638) and PTGDS (PPH4=0.619) to PC. As for the rest of the 17 circulating proteins, no significant satisfaction for assumption

H4 was revealed. However, it was worth noting that negative results did not inherently invalidate the findings obtained from MR [51].

Replication and meta-analysis

For the prominent associations found in the discovery stage, we repeated the analysis with additional data sources to inspect the robustness of conclusions. In replication phase 1, when the GWAS data for PC was obtained from a study integrating UK Biobank and GERA participants, the genetically predicted circulating level of ABO and HRG had a significant causal link with PC. Furthermore, for proteins with negative results in the primary phase, replicative proteome-wide analysis was still conducted to expand our findings. As demonstrated in Additional file1: Table S5, EVL, PCSK9, NPTX2, and GPC1 were additionally identified with potential influence on PC occurrence in replication phase 1 after multiple testing corrections. In replication phase 2, cis-pQTLs extracted from UK Biobank consortium were available for 11 out of the 21 identified plasma proteins, and 8 out of the 11 available proteins (ABO, CFD, CST5, DPT, FGF3P3, HRG, PDIA5, PTGDS) were still causally associated with PC risk. Likewise, proteome-wide MR analyses investigating 1,939 proteins from the UK Biobank Pharma Proteomics Project (UKBPPP) study were also performed in replication phase 2, whose full results can be accessed in Additional file1: Table S6. Subsequently, a meta-analysis was employed to integrate MR estimates from both the discovery and replication phases. After pooling effects and correcting P-values with FDR method, statistically significant associations with PC were observed in 11 proteins including ABO, CFD, CHST11, CHST9, CST5, ELANE, HAGH, HRG, LIPN, MAN2B2, PCSK1, PSG5, and PTGDS (Additional file1: Table S7). However, of these candidates, inconsistent causation direction between the discovery and validation phases was observed for PTGDS and MAN2B2. Accordingly, taking all the above results together, the circulating proteins were grouped into three categories (see methods). ABO and CHST11 lay in the top tier with the strongest evidence from both colocalization and replications. The tier 2 proteins incorporated proteins with successful repeats in external validation but no colocalization evidence, including CFD, ELANE, HAGH, CST5, CHST9, PSG5, PCSK1, HRG, and LIPN. And the rest of the 10 proteins were classified as tier 3 category including PTGDS, MAN2B2, DPT, PDIA5, STAMBP, FGF3P3, LINGO1, APOA5, APOM, and XXYLT1. Detailed information for protein categorizing was summarized in Table 1.

Gene expression of candidate targets and PC risk

Among the 21 identified potential therapeutic targets, eQTLs for 13 corresponding protein-coding genes in whole blood tissues were finally acquired from eQTLGen consortium. In test with two-sample MR method, the association between PC and expression levels of ELANE, HAGH, LIPN, PDIA5, PTGDS, and XXYLT1 reached a statistical significance (Additional file1: Table S8). However, it was unfortunately discovered that the effect of PTGDS gene expression was opposite to its effect at a protein level. When applying SMR method using top hit eQTL and HEIDI test as supplementary sensitivity analyses, only STAMBP was identified to correlate with PC and pass the HEIDI test. In addition, eQTL data in pancreas tissues were only available for ABO, HAGH, and MAN2B2 from GTEx v8, and after employing these eQTLs in SMR-based analysis, ABO expression yielded a positive correlation with PC (P-value=0.04), but the HEIDI test suggested a potential presence of heterogeneity (P-value for HEIDI=0.004).

Single-cell type expression in PC tissues

Single-cell type RNA sequencing data for 16 PC tissues were attracted from the dataset of GSE155698. Since ABO, PSG5, and APOA5 were not detected in the study, expressions of 18 protein-coding genes were then available for further analysis. As shown in Fig. 3A, after annotation for cell types, all cells were classified into nine clusters incorporating monocytes, T cells, epithelial cells, neutrophils, tissue stem cells, NK cells, macrophages, B cells, and endothelial cells. The cell type-specific gene expressions were demonstrated in Fig. 3B and C. Subsequently, in the Wilcoxon Rank Sum test for the differential gene expression across cell types, PTGDS was observed to be significantly enriched in tissue stem cells with FDP-corrected P-value < 0.05 and $|\log_2(\text{fold-change})| > 1$, while CFD was enriched in immune cells such as monocytes, NK cells, and neutrophils.

Pathway enrichment and protein interaction

To explore whether these identified proteins were involved in a specific biological pathway, enrichment analysis for GO terms and KEGG pathways was performed. As displayed in Fig. 4A, the top enriched pathways for biological procedure (BP) included glycoprotein biosynthetic process, chondroitin sulfate proteoglycan biosynthetic process, chondroitin sulfate biosynthetic process, and high-density lipoprotein particle assembly. In terms of cellular component (CC), these proteins were involved in cytoplasmic vesicle lumen, secretory granule lumen, triglyceride-rich plasma lipoprotein particle, and very-low-density

Table 1 Summary results of the primary MR analysis, meta-analysis, colocalization, and protein classification

Protein	Full protein name	Primary analysis			Meta-analysis results combined with replications			Colocalization			Category	
		OR (95%CI)	P-value	P _{FDR}	Sensitivity analyses	Revers effect	OR (95%CI)	P-value	P _{FDR}	PPH4		PPH3 + PPH4
ABO	Histo-blood group ABO system transferase	1.21 (1.16–1.26)	2.790E-19	4.880E-16	Passed	No	1.22 (1.18–1.25)	1.350E-40	2.835E-39	0.967	1.000	Tier 1
CHST11	Carbohydrate sulfotransferase 11	1.34 (1.18–1.51)	3.630E-06	2.118E-03	Passed	No	1.28 (1.15–1.42)	5.320E-06	9.576E-05	0.757	0.853	Tier 1
CFD	Complement factor D	1.38 (1.15–1.66)	5.099E-04	4.512E-02	Passed	No	1.33 (1.18–1.51)	4.630E-06	8.797E-05	0.052	0.398	Tier 2
ELANE	Neutrophil elastase	1.29 (1.15–1.43)	5.130E-06	2.244E-03	Passed	No	1.27 (1.14–1.41)	5.890E-06	1.001E-04	0.217	0.503	Tier 2
HAGH	Hydroxyacylglutathione hydrolase, mitochondrial	1.83 (1.36–2.47)	6.040E-05	1.230E-02	Passed	No	1.70 (1.34–2.16)	1.080E-05	1.728E-04	0.254	0.385	Tier 2
CST5	Cystatin-D	1.15 (1.06–1.25)	4.487E-04	4.512E-02	Passed	No	1.11 (1.06–1.16)	1.110E-05	1.728E-04	0.044	0.211	Tier 2
CHST9	Carbohydrate sulfotransferase 9	0.77 (0.67–0.88)	2.340E-04	3.035E-02	Passed	No	0.77 (0.68–0.87)	4.330E-05	6.062E-04	0.202	0.302	Tier 2
PSG5	Pregnancy-specific beta-1-glycoprotein 5	0.82 (0.74–0.91)	8.060E-05	1.412E-02	Passed	No	0.83 (0.75–0.91)	5.260E-05	6.838E-04	0.410	0.865	Tier 2
PCSK1	Neuroendocrine convertase 1	1.12 (1.06–1.19)	4.300E-05	1.142E-02	Passed	No	1.11 (1.05–1.16)	1.121E-04	1.345E-03	0.413	0.468	Tier 2
HRG	Histidine-rich glycoprotein	1.10 (1.04–1.160)	5.154E-04	4.512E-02	Passed	No	1.14 (1.06–1.22)	2.301E-04	2.531E-03	0.066	0.197	Tier 2
LIPN	Lipase member N	1.09 (1.04–1.14)	2.426E-04	3.035E-02	Passed	No	1.07 (1.03–1.12)	3.437E-04	3.437E-03	0.252	0.316	Tier 2
PTGDS	Prostaglandin-H2 D-isomerase	2.21 (1.56–3.12)	7.500E-06	2.626E-03	Passed	No	2.04 (1.53–2.70)	9.100E-07	1.820E-05	0.619	0.690	Tier 3
MAN2B2	Epididymis-specific alpha-mannosidase	1.13 (1.06–1.22)	5.005E-04	4.512E-02	Passed	No	1.08 (1.03–1.13)	8.781E-04	7.903E-03	0.020	0.175	Tier 3
DPT	Dermatopontin	1.23 (1.11–1.36)	1.312E-04	1.915E-02	Passed	No	1.14 (0.99–1.30)	7.252E-02	5.802E-01	0.083	0.181	Tier 3
PDIA5	Protein disulfide-isomerase A5	1.18 (1.09–1.28)	6.320E-05	1.230E-02	Passed	No	1.14 (0.96–1.35)	1.308E-01	9.156E-01	0.139	0.285	Tier 3
STAMBIP	STAM-binding protein	0.04 (0.01–0.15)	2.450E-06	2.118E-03	Passed	No	0.23 (0.03–1.75)	1.555E-01	9.330E-01	0.079	0.192	Tier 3
FGFBP3	Fibroblast growth factor-binding protein 3	1.48 (1.21–1.80)	1.062E-04	1.690E-02	Passed	No	1.26 (0.90–1.77)	1.762E-01	9.330E-01	0.414	0.788	Tier 3
LINGO1	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1	0.29 (0.16–0.53)	4.570E-05	1.142E-02	Passed	No	0.76 (0.11–5.4)	7.840E-01	1.000E+00	0.638	0.890	Tier 3
APOA5	Apolipoprotein A-V	0.85 (0.78–0.93)	3.479E-04	4.061E-02	Passed	No	0.92 (0.77–1.09)	3.356E-01	1.000E+00	0.381	0.494	Tier 3
APOM	Apolipoprotein M	0.63 (0.48–0.81)	4.772E-04	4.512E-02	Passed	No	0.82 (0.46–1.48)	5.154E-01	1.000E+00	0.472	0.585	Tier 3
XXYLT1	Xyloside xylosyltransferase 1	0.72 (0.60–0.87)	5.936E-04	4.949E-02	Passed	No	0.87 (0.58–1.31)	5.089E-01	1.000E+00	0.085	0.302	Tier 3

PTGDS and MAN2B2 were categorized into tier 3 due to inconsistent effect direction in discovery and replication phase despite the significance in meta-analysis

MR, Mendelian randomization; OR, odds ratio; CI, confidence interval; FDR, false discovery rate; PPH, posterior probabilities of hypothesis

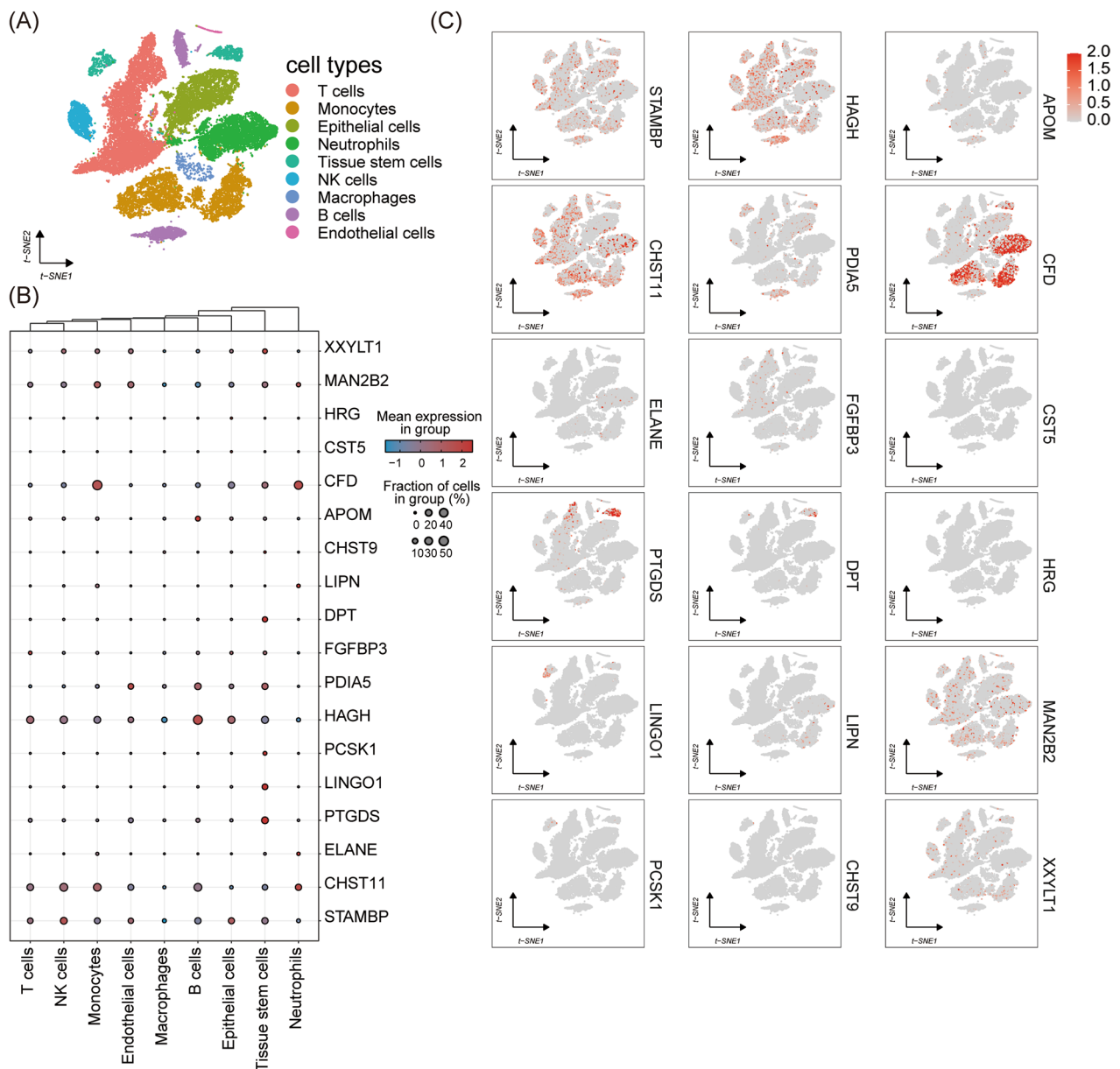


Fig. 3 Single-cell type expression in pancreatic cancer tissue for the protein-coding genes of the identified candidate targets in the discovery proteome-wide Mendelian randomization investigation. **A** represented the nine cell clusters labeled and annotated by marker genes; **B** and **C** demonstrated the expression of candidate targets in each cell type

lipoprotein particle. Moreover, for the aspect of molecular function (MF), these targets were enriched in heparin binding, glycosaminoglycan binding, sulfur compound binding, and serine-type endopeptidase activity. However, no significantly enriched KEGG pathway was highlighted. PPI network analysis was used for investigating the interplay of the identified plasma proteins. A total of 17 interaction pairs involving 14 proteins were obtained, and when setting the threshold of confidence score as 0.4, only the interaction between

APOM and APO5 (score=0.942), and interaction between APO5 and HRG (score=0.861) were observed (Fig. 4B). In addition, the mutual influence of the plasma level of these candidate proteins was inspected, and a total of 49 significant associations were identified (Fig. 4C). Of these, ABO was the one to have the greatest impact on the plasma level of other targets, and up to 8 circulating proteins (CHST11, XXYL1, MAN2B2, FGF3, PTGDS, HAGH, LINGO1, STAMBP) were up- or down-regulated by ABO level.

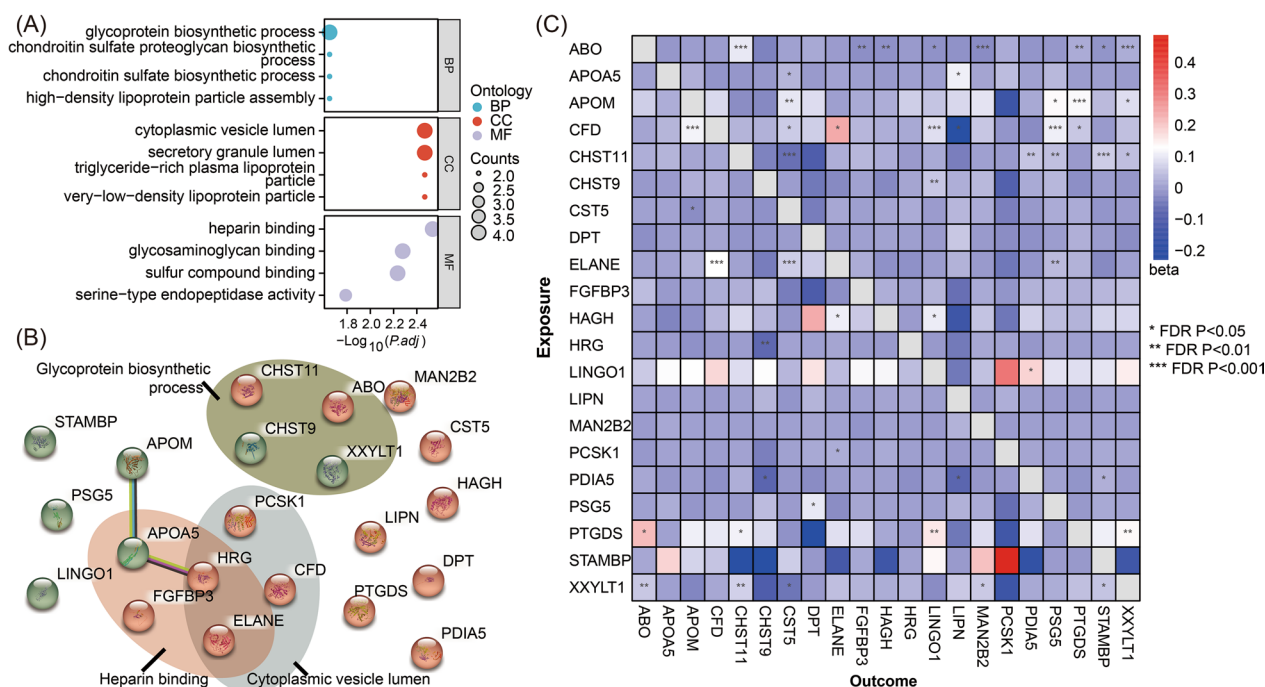


Fig. 4 The functional enrichment (A), PPI (B), and the between-protein causation analysis (C) of the identified 21 candidates in the discovery proteome-wide Mendelian randomization investigation. The red bubbles in PPI plot represent proteins acting to promote pancreatic cancer tumorigenesis, while green bubbles represent those acting to protect against pancreatic cancer. BP, biological procedure; CC, cellular component; MF, molecular function; PPI, protein-protein interaction

Single gene knock-out mouse models

The MGI resource was queried to identify the emerging phenotypes relevant to the knock-out of the potential targets. For the 21 potential target genes, among which 7 played protective roles against neoplasms, no new-onset tumor or other neoplastic trait was induced in mouse models merely by knocking out the single gene. With regard to digestive system-related traits, PCSK1 knock-out brought about chronic diarrhea and modification in intestinal goblet cells and enteroendocrine cells. Phenotypes of other systems produced by single gene knock-out could be accessed in Additional file 1: Table S9.

Mediation analysis

Alcohol consumption, body mass index (BMI), smoking, chronic pancreatitis, and type 2 diabetes mellitus were identified as risk factors for PC through previous publications [52–56]. The two-sample MR method was applied to verify the correlation between risk factors and PC. As shown in Additional file 1: Table S10, BMI (P-value < 0.001) and type 2 diabetes mellitus (P-value = 0.021) significantly increased the risk of PC. However, the impact of alcohol consumption, smoking, and chronic pancreatitis on PC occurrence was not validated in our study (P-value > 0.05). In order to inspect whether the risk factors act as mediators in protein-PC

connection, we analyzed the causal effect from the 21 candidate proteins to risk factors. After multiple testing corrections, several significant links were demonstrated: CHST11 and CHST9 were positively associated with PC; FGFBP3 and PCSK1 were negatively associated with smoking; FGFBP3 and MAN2B2 were negatively associated with type 2 diabetes mellitus while ABO was positively associated with it (Additional file1: Table S11). Considering the significance and the effect direction of above MR results, the indirect effect on PC onset was possibly valid only for ABO and CHST11, via the mediation of type 2 diabetes mellitus and BMI, respectively. The mediation effect was calculated as described in the method section, and the delta method was used to estimate the standard error. Therefore, the indirect effect of ABO on PC mediated by type 2 diabetes mellitus was 0.0049 (95%CI 0.0004–0.0094, P-value 0.035), and the corresponding mediation proportion was 2.61% (95%CI 0.19%–5.04%). The indirect effect of CHST11 on PC mediated by BMI was 0.0069 (95%CI 0.0014–0.0125, P-value = 0.015), and the corresponding mediation proportion was 2.40% (95%CI 0.47–4.33%).

Druggability assessment

In druggability assessment, eight of 21 PC-associated plasma proteins, including ELANE, PSG5, HAGH,

PCSK1, PTGDS, HRG, CHST11, and APO5, were revealed to be targeted for drug development (Additional file 1: Table S12). Drugs targeting ELANE have been applied in treatments for chronic obstructive pulmonary disorder (Alpha-1-proteinase inhibitor and Erdosteine) and neutropenia (Pegfilgrastim). Alitretinoin targeting PGS5 had been employed for topical treatment of cutaneous lesions in patients with AIDS-related Kaposi's sarcoma. Some PC-related proteins could be targeted by taking micronutrient supplements, such as Vitamin A for targeting PTGDS, Zinc chloride and Zinc sulfate for HRG, and Glutathione for HAGH. Furthermore, insulin targeting PCSK1 was widely used in treating diabetes mellitus and improving glycemic control.

Phenome-wide MR analysis

A phenome-wide MR analysis regarding the causality of target proteins on 782 non-PC traits retrieved from UK Biobank was carried out to investigate the possible side effects and repurposing. Ultimately, a total of 51 causal relationships involving 12 proteins reached statistical significance, and no association was found for APOM, CFD, CHST9, FGF3P3, HAGH, LINGO1, MAN2B2, PTGDS, and STAMB. Among the positive associations, over a half (28/51) owed to protein ABO, with the increased level of which mainly leading to thrombosis and cardiovascular events but preventing hemorrhage of digestive tract. Some other molecules also played a two-faced role in non-PC disease risk. For instance, targeting PCSK1 might benefit in alleviating cardiomegaly but elevate the risk of osteoarthritis and other arthropathies. Additionally, of these plasma proteins in relation to non-PC phenotypes, interference on APOA5, HRG, and LIPN could be repurposed as treatment for some other digestive system illness with no prominent side effect. To note, reducing the plasma concentration of PDIA5 and CST5 might increase the susceptibility of stomach cancer and uterus cancer, respectively. The Additional file 1: Table S13 documented full results for other traits influenced by PC-related plasma proteins.

Discussion

Early detection and valid treatment options for PC have long been a formidable challenge. To overcome this obstacle, increasing attention has been paid to the complex interaction of plasma proteins and cancers in recent years: on the one hand, the onset and development of tumors could be accompanied by alteration in concentration of circulating proteins due to the secretion by onco-cytes and tumor-associated stromal and immune cells, and these proteins might serve as valuable biomarkers for diagnosis and prognosis prediction [15, 57, 58]; on the other hand, these proteins are involved in multiple

process of carcinogenesis, tumor invasion, metastasis, and shaping the tumor microenvironment [59, 60]. Given the above facts, it is worthwhile to give a deep insight into the causal relationship between plasma protein and PC to assist efficient identification of potential diagnostic markers and interference targets by applying advanced analytic methodology with high-support evidence and less likelihood for confounding bias. With the emergence of the MR method, one previous research preliminarily attempted the causal inference of plasma protein levels and pan-cancer, including PC [61]. Nevertheless, the instruments used in that study were not cis-acting, increasing the risk of horizontal pleiotropy. Besides, external validation with additional data sources was not conducted, making the conclusion less persuasive. Furthermore, the role of cancer-specific risk factors in established protein-cancer correlation, as well as the possible side effects for therapies targeting these proteins were not well investigated. Hence, a more rigorous and comprehensive design was required.

In this study, after stringent quality control approaches, we obtained eligible cis-pQTLs that satisfied the basic assumptions of MR, and they were further utilized in the following proteome-wide MR analysis. In the primary proteome-wide investigation, genetically determined plasma concentration of 21 proteins was identified to significantly correlate with PC risk. Then the subsequent analyses could be summarized into three parts according to their analytic purpose. In part one analysis for validation, no reverse causation and horizontal pleiotropy were found for the 21 significant associations through bidirectional MR analysis, additional MR methods, repeated MR with varying IVs inclusion criteria, and other sensitivity analyses. The causal effect of ABO and CHST11 was confirmed by Bayesian colocalization with high-support evidence, while the effect of LINGO1 and PTGDS was confirmed with medium-support evidence. After conducting replication analysis with alternative datasets for exposure and outcomes, respectively, and after subsequent meta-analysis, the causality was no longer significant or inconsistent in direction with primary results for 10 proteins including PTGDS, MAN2B2, DPT, PDIA5, STAMB, FGF3P3, LINGO1, APOA5, APOM, and XXLT1, and they were categorized into tier 3 proteins with relatively low credibility. ABO and CHST11 lay in the tier 1 group on account of strong colocalization evidence while the rest of the other 9 proteins (CFD, ELANE, HAGH, CST5, CHST9, PGS5, PCSK1, HRG, and LIPN) were set as tier 2 category. In summary, the part one analyses successfully verified and underlined 11 plasma proteins (tier 1 and tier 2 proteins) in causal relation to PC with more convincing multi-dimension evidence. Subsequently, in part two analysis for exploration

of the underlying pathways, we tested these identified associations at a transcription level in whole blood and pancreas tissues and looked into the differential expression of these protein-coding genes in specific cell clusters. Then the enrichment analysis, PPI networks, mutual causation, and animal knock-out models were also applied attempting to comprehend the biological significance and interaction of these targets. More importantly in part two analysis, the mediation analysis indicated the partial involvement of BMI and type 2 diabetes as mediators in the procancerous effect of CHST11 and ABO on PC, respectively. Finally, in the part three analysis for druggability, these targets were prioritized by searching drug-target databases and whether the risk of other diseases would be elevated when targeting the proteins for treatment was assessed through a phenome-wide MR analysis. To sum up, drugs had been developed for eight of the candidates, among which no evidence for prominent side effects was found upon targeting HRG, HAHG, and PTGDS, suggesting their promising potentials as safe therapeutic targets.

Protein ABO and CHST11 were causally linked to PC with the most convincing evidence in the present study. ABO (Histo-blood group ABO system transferase) is a glycosyltransferase enzyme participating in the biosynthesis of A and B antigens and determining the ABO blood type of individuals. In accordance with our findings, both epidemiological and genetic evidence have revealed a decreased susceptibility of PC in O blood type individuals in comparison to non-O groups [62, 63]. PC patients PC carrying O blood group also experienced more favorable survival [64]. The exact underlying mechanisms behind this connection remain unclear and are possibly ascribed to systemic inflammatory and immune response [65–67]. Interestingly, type 2 diabetes, a known risk factor for PC, was found to partially mediate the causal effect from ABO to PC in our study, despite the controversial association between blood group and diabetes in previous reports [68, 69]. Similarly, a retrospective study revealed a higher proportion of B blood type patients among those with long-term diabetes before PC diagnosis than that among PC patients without diabetes at diagnosis [70], which also implied the intricate interaction between ABO blood type, diabetes mellitus, and PC. However, although substantial evidence has demonstrated the strong link of ABO with PC, treatment targeting ABO is elusive and challenging considering the ambiguous impact on blood type and the multitude of side effects anticipated by our phenome-wide investigation (Additional file 1: Table S13). CHST11 (Carbohydrate sulfotransferase 11) as a member of the carbohydrate sulfotransferases family, is engaged in the modification of glycosaminoglycans, specifically

in the sulfation of chondroitin sulfate, which is a crucial molecule in cancer progression and metastasis [71, 72]. And the microenvironment of PC tissue was observed enriched with chondroitin sulfate at a 22-fold increase in concentration compared to paired normal tissues [73]. It was also reported the high expression of CHST11 indicated poor prognosis for PC patients and correlated with worse clinical stage and histological grade [74]. Consistently, the pro-tumorigenesis effect of CHST11 was verified in this study, and a limited indirect effect through modulating BMI was additionally observed, throwing new insights into the possible interpretation of the association. Nevertheless, while experimental studies have shown CHST11 could drive cancer invasion, epithelial-mesenchymal transition, and cancer stem cell generation by activating signaling pathways such as Wnt/beta-catenin in multiple other types of malignancies [75, 76], its direct biological significance on PC cells has yet to be elucidated. Further laboratory evidence is required to confirm and expand our findings.

In addition to ABO and CHST11, significant correlations between another nine plasma proteins with PC were supported by external validation and meta-analysis, in spite of negative results from Bayesian colocalization. Noteworthy, there is still controversy on whether colocalization analysis violates the core MR assumptions, especially when utilizing quantitative trait loci as instruments, and negative colocalization results thereby do not necessarily attenuate the plausibility of the inferred causality [51, 77]. Of these candidate targets in tier 2, CFD, ELANE, and HAGH demonstrated the strongest statistical correlation with PC upon ordering the adjusted P-values (Table 1). CFD (Complement factor D) functions by enzymatically cleaving factor B when it is complexed with C3b in the alternative pathway of the complement system. Reports on the relationship between CFD and PC are limited, with only one retrospective study presenting that the CFD expression seemed to be irrelevant to the prognosis of PC patients undergoing neoadjuvant chemotherapy and surgery [78]. However, regarding other tumors, laboratory experiments revealed that CFD stimulated proliferation in cutaneous squamous cell carcinoma by modulating ERK1/2 signaling pathway [79]. In particular, CFD is also known as one of the obesity-driven biomarkers, and CFD along with its downstream effector hepatocyte growth factor secreted by adipocytes could augment the properties of cancer stem cells in breast cancer [80, 81]. Given that PC is another malignancy linked to obesity and adipose accumulation, future investigation into the role of CFD within the mechanisms through which obesity contributes to PC could be of vital value. ELANE (Neutrophil elastase) is a serine protease predominantly secreted by neutrophils and is implicated

in the degradation of extracellular matrix proteins in the process of inflammation against pathogens [82]. It is also one of the essential components of neutrophil extracellular traps, which has been proven to activate pancreatic stellate cells to form a thick, fibrotic stroma and accelerate PC growth [83]. A previous report illuminated that ELANE played a mediator role in intratumoral bacteria and PC carcinogenesis, by shaping a pro-inflammatory tumor microenvironment [84]. In line with these findings, an elevated concentration of ELANE at a plasma protein or transcription level led to increased PC risk in our study. Besides, the declined secretion of tumor necrosis factor was highlighted after ELANE knock-out in mouse models. And drugs targeting ELANE have been approved in treating some inflammatory diseases, such as chronic obstructive pulmonary disorder. HAGH (Hydroxyacylglutathione hydrolase), also described as Glyoxalase II (GLO2), together with Glyoxalase I (GLO1) constitutes the glyoxalase system that is involved in the detoxification of methylglyoxal produced during the glycolytic pathway. In this study, the genetically determined plasma concentration of HAGH/GLO2 was positively associated with PC susceptibility. Consistently, numerous researches have implied the involvement of GLO1 and GLO2 in progression of multiple tumors. In PC, up-regulation of GLO1 was spotted in cancerous tissues and indicated poor outcome and acquired resistance to gemcitabine [85, 86]. In comparison, studies on GLO2 are scant and primarily focused on urological malignancies. For instance, GLO2 was observed to promote proliferation and elude apoptosis via mechanisms involving p53-p21 axis [87].

This study has a number of noticeable advantages. To the best of our knowledge, the current study presented the most extensive and comprehensive proteome-wide MR analysis for PC. The breadth, depth, and rigorosity of this study allowed for a more robust identification of promising targets for the development of screening biomarkers and therapeutic drugs for PC. First, numerous measures were taken to evade violation of basic MR assumptions and diminish the risk of confounding bias. We excluded pQTLs located within the MHC region or distant from the vicinity of the corresponding protein-coding gene (trans-pQTL). Considering MR analysis is susceptible to the IVs selection, we repeated our analysis with additional IVs inclusion criteria that has been employed in previous proteome-wide MR studies [88, 89]. A series of supplementary sensitivity analyses, reverse MR analysis, and colocalization analysis were conducted to enhance the robustness of identified associations. Consistent results from replicative MR analysis with replaced data source and subsequent meta-analysis, as well as MR or SMR using eQTL data of blood and

pancreas tissues, also minimized the false positive risk of the conclusions. Second, evidence from enrichment analysis, PPI, mutual causation analysis, and single gene knock-out models, could provide potential views on how these candidate targets interact with PC. Third, we clarified a partial involvement of established PC risk factors, especially type 2 diabetes and BMI, in the pathogenic pathway of these plasma proteins. Last but not least, these identified targets were prioritized with druggability and side effects with three distinct drug-target databases and a phenome-wide MR investigation.

However, there are still several limitations. First, the enrolled populations in this study were mostly European individuals, and this restricted the expansion of our conclusions to other ancestries. Second, although employing cis-pQTLs instead of trans-pQTLs as instruments could avoid horizontal pleiotropy as much as possible, it would reduce the number of assessable candidates as a result of no eligible SNP for some proteins. To counteract this, we included up to six distinct proteomic studies in the primary stage, but it was still noteworthy that measurement bias might exist among these researches. Third, this study principally focused on the proteins in plasma, and their effects in pancreas tissues were not explored due to insufficient available data. Similarly, because of the lack of eligible eQTLs, investigation regarding the association at an expression level was unavailable for some of the candidate targets. Moreover, it was worth noting that the STRING database applied for PPI analysis was not PC tissues-specific, meaning these results should be interpreted more conservatively. Last, despite the superiority of MR approach in causality inference, it was rarely possible to thoroughly eliminate confounding bias or reverse causation. Consequently, large epidemiological and experimental studies are warranted to support the above results, and our plans are currently underway to gradually carry out cell experiments and animal experiments to offer further evidence for the pathogenic role of part of interested proteins.

In conclusion, a total of 21 plasma proteins were identified with etiological significance for PC. Two and nine of them were prioritized with the most convincing and medium-support evidence, respectively. With future effective validation, these candidate proteins might serve as novel biomarkers in PC early detection and promising druggable targets for PC treatment.

Abbreviations

PC	Pancreatic cancer
BMI	Body mass index
pQTL	Protein quantitative trait loci
Cis-pQTL	Cis-acting protein quantitative trait loci
MHC	Major histocompatibility complex
MAF	Minor allele frequency
eQTL	Expression quantitative trait loci

GWAS	Genome-wide association study
UKBPPP	UK Biobank Pharma Proteomics Project
MR	Mendelian randomization
SMR	Summary-data-based Mendelian randomization
HEIDI	Heterogeneity in independent instrument
LD	Linkage disequilibrium
IVs	Instrumental variables
SNP	Single nucleotide polymorphism
PPI	Protein–protein interaction
KEGG	Kyoto Encyclopedia of Genes and Genomes
STRING	Search Tool for the Retrieval of Interacting Genes
MGI	Mouse Genome Informatics
OR	Odds ratio
CI	Confidence interval
GEO	Gene Expression Omnibus
scRNA-seq	Single-cell RNA sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05363-9>.

Additional file 1 (XLSX 4422 KB)

Acknowledgements

We appreciate the participants and investigators of the FinnGen consortium, UK Biobank, and all other proteomic and GWAS studies involved in the present research.

Author contributions

SZ, DF, and JL designed the study. BT, YG, and SZ were responsible for manuscript writing. JG, HL, CZ, and ST participated in the data acquisition and processing. SZ and SJ carried out data analysis and results interpretation. SJ, DF, and JL reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Chinesisch-Deutsche Kooperationsgruppe: Precision Medicine in Pancreatic Cancer (GZ1456) and Shanghai Science and Technology Innovation Action Plan (22511106203).

Availability of data and materials

The data of FinnGen can be accessed at <https://www.finnngen.fi/en>. The data of UK Biobank can be accessed at <https://pheweb.org/UKB-SAIGE/>. The GWAS summary statistics for PC based on a meta-analysis of the UK Biobank and GERA cohorts can be accessed at https://github.com/Wittelab/pancancer_pleiotropy. GWAS summary statistics for PC risk factors are available at UK Biobank (<https://pheweb.org/UKB-SAIGE/>) and IEU Open GWAS Project (<https://gwas.mrcieu.ac.uk/>). The pQTL data can be accessed at <https://www.decode.com/summarydata/>, <https://www.synapse.org/#!Synapse:syn51364943/files/>, <https://gwas.mrcieu.ac.uk/>, <https://www.ebi.ac.uk/gwas/publications/37563310>, and <https://www.ebi.ac.uk/gwas/publications/33067605>. The expression (eQTL) data for whole blood and pancreas tissue can be accessed at <https://eqtlgen.org/phase1.html> and https://yanglab.westlake.edu.cn/data/SMR/GTEx_v8_cis_eqtl_summary.html. The single-cell RNA sequencing data of PC tissue can be accessed at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155698>.

Declarations

Ethics approval and consent to participate

All studies were previously approved by respective institutional review boards (IRBs). No new IRB approval was required.

Consent for publications

Not applicable.

Competing interests

None.

Author details

¹Department of Pancreatic Surgery, Huashan Hospital, Fudan University, Shanghai 200040, China. ²School of Medicine, Fudan University, Shanghai 200240, China. ³State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China.

Received: 22 March 2024 Accepted: 29 May 2024

Published online: 10 June 2024

References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70:7–30.
- Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. *World J Oncol*. 2019;10:10–27.
- Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. *Lancet*. 2020;395:2008–20.
- Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med*. 2021;4:65.
- Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat Rev Gastroenterol Hepatol*. 2021;18:493–502.
- Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8:14357.
- Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1:845–67.
- Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet*. 2021;22:19–37.
- Saarikoski LA, Huupponen RK, Viikari JS, Marniemi J, Juonala M, Kähönen M, Raitakari OT. Adiponectin is related with carotid artery intima-media thickness and brachial flow-mediated dilatation in young adults—the Cardiovascular Risk in Young Finns study. *Ann Med*. 2010;42:603–11.
- Oikonen M, Wendelin-Saarenhovi M, Siitonen N, Sainio A, Juonala M, Kähönen M, Lyytikäinen LP, Seppälä I, Lehtimäki T, Viikari JS, et al. Tissue inhibitor of matrix metalloproteinases 4 (TIMP4) in a population of young adults: relations to cardiovascular risk markers and carotid artery intima-media thickness. The cardiovascular risk in young Finns study. *Scand J Clin Lab Invest*. 2012;72:540–6.
- Du Clos TW, Mold C. C-reactive protein: an activator of innate immunity and a modulator of adaptive immunity. *Immunol Res*. 2004;30:261–77.
- Bhardwaj M, Weigl K, Tikik K, Holland-Letz T, Schrotz-King P, Borchers CH, Brenner H. Multiplex quantitation of 270 plasma protein markers to identify a signature for early detection of colorectal cancer. *Eur J Cancer*. 2020;127:30–40.
- Landegren U, Hammond M. Cancer diagnostics based on plasma protein biomarkers: hard times but great expectations. *Mol Oncol*. 2021;15:1715–26.
- Enroth S, Berggrund M, Lycke M, Broberg J, Lundberg M, Assarsson E, Olovsson M, Ståhlberg K, Sundfeldt K, Gyllensten U. High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Commun Biol*. 2019;2:221.
- Davies MPA, Sato T, Ashoor H, Hou L, Lilioglou T, Yang R, Field JK. Plasma protein biomarkers for early prediction of lung cancer. *EBioMedicine*. 2023;93: 104686.
- Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007;166:646–55.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1–22.
- Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005;366:1906–8.
- Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27:3253–65.

20. Jiang Z, Mou Y, Wang H, Li L, Jin T, Wang H, Liu M, Jin W. Causal effect between gut microbiota and pancreatic cancer: a two-sample Mendelian randomization study. *BMC Cancer*. 2023;23:1091.
21. Sun R, Xu H, Liu F, Zhou B, Li M, Sun X. Unveiling the intricate causal nexus between pancreatic cancer and peripheral metabolites through a comprehensive bidirectional two-sample Mendelian randomization analysis. *Front Mol Biosci*. 2023;10:1279157.
22. Zhong H, Liu S, Zhu J, Xu TH, Yu H, Wu L. Elucidating the role of blood metabolites on pancreatic cancer risk using two-sample Mendelian randomization analysis. *Int J Cancer*. 2024;154:852–62.
23. Mälärstig A, Grassmann F, Dahl L, Dimitriou M, McLeod D, Gabrielson M, Smith-Byrne K, Thomas CE, Huang TH, Forsberg SKG, et al. Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. *Nat Commun*. 2023;14:7680.
24. Sun J, Zhao J, Jiang F, Wang L, Xiao Q, Han F, Chen J, Yuan S, Wei J, Larsson SC, et al. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. *Genome Med*. 2023;15:75.
25. Cai YX, Wu YQ, Liu J, Pan H, Deng W, Sun W, Xie C, Huang XF. Proteome-wide analysis reveals potential therapeutic targets for colorectal cancer: a two-sample Mendelian randomization study. *BMC Cancer*. 2023;23:1188.
26. Wu Y, Wang Z, Yang Y, Han C, Wang L, Kang K, Zhao A. Exploration of potential novel drug targets and biomarkers for small cell lung cancer by plasma proteome screening. *Front Pharmacol*. 2023;14:1266782.
27. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, Reeve MP, Laivuori H, Aavikko M, Kaunisto MA, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613:508–18.
28. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismisdottir EL, Gunnarsdottir K, Helgason A, Oddsson A, Halldorsson BV, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. 2021;53:1712–21.
29. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73–9.
30. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Fränberg M, Sennblad B, Baldassarre D, Veglia F, Humphries SE, Rauramaa R, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet*. 2017;13:e1006706.
31. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, Page K, Zhernakova DV, Wu Y, Peters J, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab*. 2020;2:1135–48.
32. !!! INVALID CITATION !!!
33. Papanimitriou N, Dimou N, Tsilidis KK, Banbury B, Martin RM, Lewis SJ, Kazmi N, Robinson TM, Albanes D, Aleksandrova K, et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nat Commun*. 2020;11:597.
34. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018. <https://doi.org/10.7554/eLife.34408>.
35. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35:1880–906.
36. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44:512–25.
37. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693–8.
38. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10:e1004383.
39. Chen J, Xu F, Ruan X, Sun J, Zhang Y, Zhang H, Zhao J, Zheng J, Larsson SC, Wang X, et al. Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. *EBioMedicine*. 2023;89: 104494.
40. Kia DA, Zhang D, Guelfi S, Manzoni C, Hubbard L, Reynolds RH, Botía J, Rytén M, Ferrari R, Lewis PA, et al. Identification of candidate parkinson disease genes by integrating genome-wide association study, expression, and epigenetic data sets. *JAMA Neurol*. 2021;78:464–72.
41. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*. 2023;622:329–38.
42. Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeeff SE, Blatchins MA, Cavazos TB, Corley DA, Emami NC, Hoffman JD, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun*. 2020;11:4423.
43. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, Gluud C, Devereaux PJ, Wetterslev J. Evolution of heterogeneity (I²) estimates and their 95% confidence intervals in large meta-analyses. *PLoS ONE*. 2012;7:e39471.
44. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53:1300–10.
45. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48:481–7.
46. Moncada R, Barkley D, Wagner F, Chiodini M, Devlin JC, Baron M, Hajdu CH, Simeone DM, Yanai I. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol*. 2020;38:333–42.
47. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36:411–20.
48. Liu J, Yuan Q, Ren J, Li Y, Zhang Y, Shang D. Single-cell sequencing and bulk RNA sequencing reveal a cell differentiation-related multigene panel to predict the prognosis and immunotherapy response of hepatocellular carcinoma. *Chin Med J*. 2023;136:485–7.
49. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4:P3.
50. Yuan J, Xiong X, Zhang B, Feng Q, Zhang J, Wang W, Tang J. Genetically predicted C-reactive protein mediates the association between rheumatoid arthritis and atlantoaxial subluxation. *Front Endocrinol*. 2022;13:1054206.
51. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, Wallace C, Burgess S. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am J Hum Genet*. 2022;109:767–82.
52. Park SM, Kim KB, Han JH, Kim N, Kang TU, Swan H, Kim HJ. Incidence and risk of pancreatic cancer in patients with acute or chronic pancreatitis: a population-based cohort study. *Sci Rep*. 2023;13:18930.
53. Arjani S, Saint-Maurice PF, Julián-Serrano S, Eibl G, Stolzenberg-Solomon R. Body mass index trajectories across the adult life course and pancreatic cancer risk. *JNCI Cancer Spectr*. 2022. <https://doi.org/10.1093/jncics/pkac066>.
54. Védie AL, Laouali N, Gelot A, Severi G, Boutron-Ruault MC, Rebours V. Childhood and adulthood passive and active smoking, and the ABO group as risk factors for pancreatic cancer in women. *United Eur Gastroenterol J*. 2023. <https://doi.org/10.1002/ueg2.12487>.
55. Okita Y, Sobue T, Zha L, Kitamura T, Iwasaki M, Inoue M, Yamaji T, Tsugane S, Sawada N. Association between alcohol consumption and risk of pancreatic cancer: the Japan public health center-based prospective study. *Cancer Epidemiol Biomarkers Prev*. 2022;31:2011–9.
56. Jensen MH, Cichosz SL, Hejlesen O, Henriksen SD, Drewes AM, Olesen SS. Risk of pancreatic cancer in people with new-onset diabetes: a Danish nationwide population-based cohort study. *Pancreatol*. 2023;23:642–9.
57. Christensen TD, Maag E, Larsen O, Feltøft CL, Nielsen KR, Jensen LH, Leerhøj B, Hansen CP, Chen IM, Nielsen DL, Johansen JS. Development and validation of circulating protein signatures as diagnostic biomarkers for biliary tract cancer. *JHEP Rep*. 2023;5: 100648.
58. Grassmann F, Mälärstig A, Dahl L, Bendes A, Dale M, Thomas CE, Gabrielson M, Hedman ÅK, Eriksson M, Margolin S, et al. The impact of circulating protein levels identified by affinity proteomics on short-term, overall breast cancer risk. *Br J Cancer*. 2024;130:620–7.

59. Dagnino S, Bodinier B, Guida F, Smith-Byrne K, Petrovic D, Whitaker MD, Haugdahl Nøst T, Agnoli C, Palli D, Sacerdote C, et al. Prospective identification of elevated circulating CDCP1 in patients years before onset of lung cancer. *Cancer Res.* 2021;81:3738–48.
60. Cohen A, Wang E, Chisholm KA, Kostyleva R, O'Connor-McCourt M, Pinto DM. A mass spectrometry-based plasma protein panel targeting the tumor microenvironment in patients with breast cancer. *J Proteomics.* 2013;81:135–47.
61. Sun J, Luo J, Jiang F, Zhao J, Zhou S, Wang L, Zhang D, Ding Y, Li X. Exploring the cross-cancer effect of circulating proteins and discovering potential intervention targets for 13 site-specific cancers. *J Natl Cancer Inst.* 2023. <https://doi.org/10.1093/jnci/djad247>.
62. Liu M, Ji S, Xu W, Liu W, Qin Y, Xiang J, Hu Q, Sun Q, Zhang Z, Xu X, Yu X. ABO blood group and the risk of pancreatic neoplasms in chinese han population: a study at shanghai pancreatic cancer institute. *Pancreas.* 2019;48:e65–6.
63. Lennon AM, Klein AP, Goggins M. ABO blood group and other genetic variants associated with pancreatic cancer. *Genome Med.* 2010;2:39.
64. Tezuka K, Ohgi K, Okamura Y, Sugiura T, Ito T, Yamamoto Y, Ashida R, Otsuka S, Todaka A, Uesaka K. The prognostic impact of ABO blood type in pancreatic cancer: relevance to adjuvant chemotherapy. *J Hepatobiliary Pancreat Sci.* 2022;29:922–31.
65. Paré G, Chasman DI, Kellogg M, Zee RY, Rifai N, Badola S, Miletich JP, Ridker PM. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6578 women. *PLoS Genet.* 2008;4: e1000118.
66. Liumbruno GM, Franchini M. Beyond immunohaematology: the role of the ABO blood group in human diseases. *Blood Transfus.* 2013;11:491–9.
67. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, Bueno-de-Mesquita HB, Gross M, Helzlsouer K, Jacobs EJ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet.* 2009;41:986–90.
68. Fagherazzi G, Gusto G, Clavel-Chapelon F, Balkau B, Bonnet F. ABO and rhesus blood groups and risk of type 2 diabetes: evidence from the large E3N cohort study. *Diabetologia.* 2015;58:519–22.
69. Cano EA, Esguerra MA, Batausa AM, Baluyut JR, Cadiz R, Docto HF, Encabo JR, Gomez RM, Sadang MG. Association between ABO blood groups and type 2 diabetes mellitus: a meta-analysis. *Curr Diabetes Rev.* 2023;19: e270422204139.
70. Egawa N, Lin Y, Tabata T, Kuruma S, Hara S, Kubota K, Kamisawa T. ABO blood type, long-standing diabetes, and the risk of pancreatic cancer. *World J Gastroenterol.* 2013;19:2537–42.
71. Theocharis AD, Tsolakis I, Tzanakakis GN, Karamanos NK. Chondroitin sulfate as a key molecule in the development of atherosclerosis and cancer progression. *Adv Pharmacol.* 2006;53:281–95.
72. Willis CM, Klüppel M. Chondroitin sulfate-E is a negative regulator of a pro-tumorigenic Wnt/beta-catenin-Collagen 1 axis in breast cancer cells. *PLoS ONE.* 2014;9: e103966.
73. Theocharis AD, Tsara ME, Papageorgacopoulou N, Karavias DD, Theocharis DA. Pancreatic carcinoma is characterized by elevated content of hyaluronan and chondroitin sulfate with altered disaccharide composition. *Biochim Biophys Acta.* 2000;1502:201–6.
74. Zhang P, Chen D, Cui H, Luo Q. High expression of CHST11 correlates with poor prognosis and tumor immune infiltration of pancreatic cancer. *Clin Lab.* 2022. <https://doi.org/10.7754/Clin.Lab.2022.211239>.
75. Behrens A, Jousheghany F, Yao-Borengasser A, Siegel ER, Kieber-Emmons T, Monzavi-Karbassi B. Carbohydrate (chondroitin 4) sulfotransferase-11-mediated induction of epithelial-mesenchymal transition and generation of cancer stem cells. *Pharmacology.* 2020;105:246–59.
76. Chang WM, Li LJ, Chiu IA, Lai TC, Chang YC, Tsai HF, Yang CJ, Huang MS, Su CY, Lai TL, et al. The aberrant cancer metabolic gene carbohydrate sulfotransferase 11 promotes non-small cell lung cancer cell metastasis via dysregulation of ceruloplasmin and intracellular iron balance. *Transl Oncol.* 2022;25: 101508.
77. Yin Q, Zhu L. Does co-localization analysis reinforce the results of Mendelian randomization? *Brain.* 2024;147:e7–8.
78. Sahní S, Nahm C, Ahadi MS, Sioson L, Byeon S, Chou A, Maloney S, Moon E, Pavlakis N, Gill AJ, et al. Gene expression profiling of pancreatic ductal adenocarcinomas in response to neoadjuvant chemotherapy. *Cancer Med.* 2023;12:18050–61.
79. Rahmati Nezhad P, Riihilä P, Knuutila JS, Viikklepp K, Peltonen S, Kallajoki M, Meri S, Nissinen L, Kähäri VM. Complement factor D is a novel biomarker and putative therapeutic target in cutaneous squamous cell carcinoma. *Cancers.* 2022;14:305.
80. Mizuno M, Khaledian B, Maeda M, Hayashi T, Mizuno S, Munetsuna E, Watanabe T, Kono S, Okada S, Suzuki M, et al. Adipsin-dependent secretion of hepatocyte growth factor regulates the adipocyte-cancer stem cell interaction. *Cancers.* 2021;13:4238.
81. Goto H, Shimono Y, Funakoshi Y, Imamura Y, Toyoda M, Kiyota N, Kono S, Takao S, Mukohara T, Minami H. Adipose-derived stem cells enhance human breast cancer growth and cancer stem cell-like properties through adipsin. *Oncogene.* 2019;38:767–79.
82. Standish AJ, Weiser JN. Human neutrophils kill *Streptococcus pneumoniae* via serine proteases. *J Immunol.* 2009;183:2602–9.
83. Miller-Ocuin JL, Liang X, Boone BA, Doerfler WR, Singhi AD, Tang D, Kang R, Lotze MT, Zeh HJ 3rd. DNA released from neutrophil extracellular traps (NETs) activates pancreatic stellate cells and enhances pancreatic tumor growth. *Oncoimmunology.* 2019;8: e1605822.
84. Tan Q, Ma X, Yang B, Liu Y, Xie Y, Wang X, Yuan W, Ma J. Periodontitis pathogen *Porphyromonas gingivalis* promotes pancreatic tumorigenesis via neutrophil elastase from tumor-associated neutrophils. *Gut Microbes.* 2022;14:2073785.
85. Crake R, Gasmi I, Dehaye J, Lardinois F, Peiffer R, Maloujahmoum N, Agirman F, Koopmansch B, D'Haene N, Azurmendi Senar O, et al. Resistance to gemcitabine in pancreatic cancer is connected to methylglyoxal stress and heat shock response. *Cells.* 2023;12:1214.
86. Wang Y, Kuramitsu Y, Ueno T, Suzuki N, Yoshino S, Iizuka N, Akada J, Kitagawa T, Oka M, Nakamura K. Glyoxalase I (GLO1) is up-regulated in pancreatic cancerous tissues compared with related non-cancerous tissues. *Anticancer Res.* 2012;32:3219–22.
87. Antognelli C, Ferri I, Bellezza G, Siccu P, Love HD, Talesa VN, Sidoni A. Glyoxalase 2 drives tumorigenesis in human prostate cells in a mechanism involving androgen receptor and p53–p21 axis. *Mol Carcinog.* 2017;56:2112–26.
88. Bourgault J, Abner E, Manikpurage HD, Pujol-Gualdo N, Laisk T, Gobeil É, Gagnon E, Girard A, Mitchell PL, Thériault S, et al. Proteome-wide Mendelian randomization identifies causal links between blood proteins and acute pancreatitis. *Gastroenterology.* 2023;164:953–965.e953.
89. Wang X, Huang T, Jia J. Proteome-wide Mendelian randomization analysis identified potential drug targets for atrial fibrillation. *J Am Heart Assoc.* 2023;12: e029003.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.